Entropy and Index of Coincidence, lower bounds

Flemming Topsøe *
University of Copenhagen

Abstract

Second order lower bounds for the entropy function (H) expressed in terms of the index of coincidence (IC) are derived. Equivalently, these bounds involve entropy and Rényi entropy of order 2 (H_2). The constants found either explicitly or implicitly are best possible in a natural sense.

Keywords Entropy, index of coincidence, Rényi entropy, measure of roughness.

1 Background, introduction

We study probability distributions over the natural numbers. The set of all such distributions is denoted $M_+^1(\mathbb{N})$ and the set of $P \in M_+^1(\mathbb{N})$ which are supported by $\{1, 2, \cdots, n\}$ is denoted $M_+^1(n)$.

We use U_k to denote a generic uniform distribution over a k-element set, and if also $U_{k+1}, U_{k+2}, \cdots$ are considered, it is assumed that the supports are increasing. By H and by IC we denote *entropy* and *index of coincidence*, respectively, i.e.

$$H(P) = -\sum_{k=1}^{\infty} p_k \, ln \, p_k \, ,$$

$$IC(P) = \sum_{k=1}^{\infty} p_k^2 \, .$$

*This work was supported by the Danish Natural Science Research Council and by INTAS, project 00-738.

Results involving index of coincidence may be reformulated in terms of *Rényi entropy of order* 2 (H_2) as

$$H_2(P) = -\ln IC(P).$$

In Harremoës and Topsøe [4] the exact range of the map $P \curvearrowright (IC(P), H(P))$ with P varying over either $M_+^1(n)$ or $M_+^1(\mathbb{N})$ was determined. Earlier related work includes Kovalevskij [6], Tebbe and Dwyer [7], Ben-Bassat [1], Golic [3] and Feder and Merhav [2]. The ranges in question, termed *IC/H-diagrams*, were denoted Δ, respectively Δ_n:

$$\Delta = \{(IC(P), H(P)) \mid P \in M_+^1(\mathbb{N})\},$$
$$\Delta_n = \{(IC(P), H(P)) \mid P \in M_+^1(n)\}.$$

By Jensen's inequality we find that $H(P) \geq -\ln IC(P)$, thus the logarithmic curve $t \curvearrowright (t, -\ln t)$; $0 < t \leq 1$ is a lower bounding curve for the IC/H-diagrams. The points $Q_k = \left(\frac{1}{k}, \ln k\right)$; $k \geq 1$ all lie on this curve. They correspond to the uniform distributions: $(IC(U_k), H(U_k)) = (\frac{1}{k}, \ln k)$. No other points in the diagram Δ lie on the logarithmic curve, in fact, Q_k; $k \geq 1$ are extremal points of Δ in the sense that the convex hull they determine contains Δ. No smaller set has this property.

Figure 1, adapted from [4], illustrates the situation for the restricted diagrams Δ_n. The key result of [4] states that Δ_n is the bounded region determinated by a certain Jordan curve determined by n smooth arcs, viz. the "upper arc" connecting Q_1 and Q_n and then $n-1$ "lower arcs" connecting Q_n with Q_{n-1}, Q_{n-1} with Q_{n-2} etc. until Q_2 which is connected with Q_1.

In [4], see also [8], the main result was used to develop concrete upper bounds for the entropy function. Our concern here will be lower bounds. The study depends crucially on the nature of the lower arcs. In [4] these arcs were identified. Indeed, the arc connecting Q_{k+1} with Q_k is the curve which may be parametrized as follows:

$$s \curvearrowright \vec{\varphi}((1-s)U_{k+1} + s\,U_k)$$

with s running through the unit interval and with $\vec{\varphi}$ denoting the *IC/H-map* given by $\vec{\varphi}(P) = (IC(P), H(P))$; $P \in M_+^1(\mathbb{N})$.[1]

The distributions in $M_+^1(\mathbb{N})$ fall in *IC-complexity classes*. The k'th class consists of all $P \in M_+^1(\mathbb{N})$ for which $IC(U_{k+1}) < IC(P) \leq IC(U_k)$ or, equivalently, for which $\frac{1}{k+1} < IC(P) \leq \frac{1}{k}$. In order to determine good lower bounds for the entropy of a distribution P, one first determines the IC-complexity class k of P. One then determines that value of $s \in]0, 1]$ for

[1]In passing we note that $s \curvearrowright \vec{\varphi}((1-s)U_1 + sU_n)$ parametrizes the upper arc.

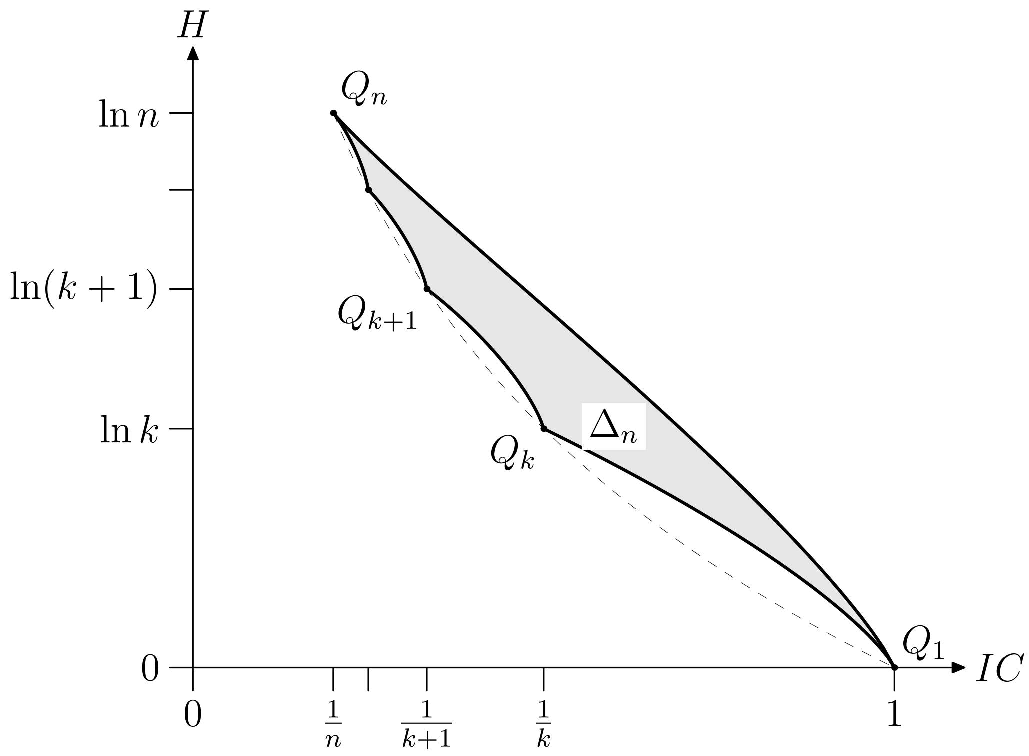

Figure 1: The restricted IC/H-diagram Δ_n, $(n = 5)$.

which $IC(P_s) = IC(P)$ with $P_s = (1-s)U_{k+1} + s\,U_k$. Then $H(P) \geq H(P_s)$ is the theoretically best lower bound of $H(P)$ in terms of $IC(P)$.

In order to write the sought lower bounds for $H(P)$ in a convenient form, we introduce the k'th *relative measure of roughness* by

$$\overline{MR}_k(P) = \frac{IC(P) - IC(U_{k+1})}{IC(U_k) - IC(U_{k+1})} = k(k+1)\left(IC(P) - \frac{1}{k+1}\right). \quad (1)$$

This definition applies to any $P \in M_+^1(\mathbb{N})$ but really, only distributions of IC-complexity class k will be of relevance to us. Clearly, $\overline{MR}_k(U_{k+1}) = 0$, $\overline{MR}_k(U_k) = 1$ and for any distribution of IC-complexity class k, $0 \leq \overline{MR}_k(P) \leq 1$. A simple calculation reveals the fact that for a distribution on the lower arc connecting Q_{k+1} with Q_k the following useful structural relation holds:

$$\overline{MR}_k((1-s)U_{k+1} + s\,U_k) = s^2\,. \quad (2)$$

In view of the above said, it follows that for any distribution P of IC-complexity class k, the theoretically best lower bound for $H(P)$ in terms of $IC(P)$ is given by the inequality

$$H(P) \geq H\big((1-x)U_{k+1} + x\,U_k\big) \quad (3)$$

where x is determined so that P and $(1-x)U_{k+1} + x\,U_k$ have the same index of coincidence, i.e.

$$x^2 = \overline{MR}_k(P)\,. \quad (4)$$

By writing out the right-hand-side of (3) we then obtain the best lower bound of the type discussed. Doing so one obtains a quantity of mixed type, involving logarithmic and rational functions. It is desirable to search for structurally simpler bounds, getting rid of logarithmic terms. The simplest and possibly most useful bound of this type is the linear bound

$$H(P) \geq H(U_k)\overline{MR}_k(P) + H(U_{k+1})(1 - \overline{MR}_k(P)) \quad (5)$$

which expresses the fact mentioned previously regarding the extremal position of the points Q_k in relation to the set Δ. Note that (5) is the best linear lower bound as equality holds for $P = U_{k+1}$ as well as for $P = U_k$. Another comment is that though (5) was developed with a view to distributions of IC-complexity class k, the inequality holds for all $P \in M_+^1(\mathbb{N})$ (but is weaker than the trivial bound $H \geq -\,ln\,IC$ for distributions of other IC-complexity classes).

Writing (5) directly in terms of $IC(P)$ we obtain the inequalities

$$H(P) \geq \alpha_k - \beta_k\,IC(P)\,;\ k \geq 1 \quad (6)$$

with α_k and β_k given via the constants

$$u_k = \ln{(1+\frac{1}{k})^k} = k \ln{(1+\frac{1}{k})} \tag{7}$$

by

$$\alpha_k = \ln{(k+1)} + u_k\,, \tag{8}$$
$$\beta_k = (k+1)u_k\,. \tag{9}$$

Note that $u_k \uparrow 1$. [2]

In the present paper we shall develop sharper inequalitites than those above by adding a second order term. More precisely, for $k \geq 1$, we denote by γ_k the largest constant such that the inequality

$$H \geq \ln k\, \overline{MR}_k + \ln{(k+1)}\,(1-\overline{MR}_k) + \frac{\gamma_k}{2k}\overline{MR}_k(1-\overline{MR}_k) \tag{10}$$

holds for all $P \in M_+^1(\mathbb{N})$. Here, $H = H(P)$ and $\overline{MR}_k = \overline{MR}_k(P)$. Expressed directly in terms of $IC = IC(P)$, (10) states that

$$H \geq \alpha_k - \beta_k\, IC + \frac{\gamma_k}{2}k(k+1)^2\left(IC - \frac{1}{k+1}\right)\left(\frac{1}{k} - IC\right) \tag{11}$$

for $P \in M_+^1(\mathbb{N})$.

The basic results of our paper may be summarized as follows:

The constants $(\gamma_k)_{k\geq 1}$ increase with $\gamma_1 = \ln 4 - 1 \approx 0.3863$ and with limit value $\gamma \approx 0.9640$.

More substance will be given to this result by developing rather narrow bounds for the γ_k's in terms of γ and by other means.

The refined second order inequalities are here published in their own right. The authors motivation to develop these bounds lies in applications to problems of exact prediction in Bernoulli models. See the discussion in Section 3.

[2]Concrete algebraic bounds for the u_k, which, via (6), may be used to obtain concrete lower bounds for $H(P)$, are given by

$$\frac{2k}{2k+1} \leq u_k \leq \frac{2k+1}{2k+2}\,.$$

This follows directly from (6) of [9] (as $u_k = \lambda(\frac{1}{k})$ in the notation of that manuscript).

2 Basic results

The key to our results is the inequality (3) with x determined by (4) [3]. This leads to the following analytical expression of γ_k:

Lemma 1. *For $k \geq 1$ define $f_k : [0,1] \to [0,\infty]$ by*

$$f_k(x) = \frac{2k}{x^2(1-x^2)}\left[-\frac{k+x}{k+1}\, ln\, \left(1+\frac{x}{k}\right) - \frac{1-x}{k+1}\, ln\,(1-x) + x^2\, ln\, \left(1+\frac{1}{k}\right)\right].$$

Then $\gamma_k = \inf\{f_k(x) \mid 0 < x < 1\}$.

Proof. By the defining relation (10) and by (3) with x given by (4), recalling also the relation (2), we realize that γ_k is the infimum over $x \in]0,1[$ of

$$\frac{2k}{x^2(1-x^2)}\left[H\left((1-x)U_{k+1} + xU_k\right) - \, ln\, k \cdot x^2 - \, ln\,(k+1)\cdot(1-x^2)\right].$$

Writing out the entropy of $(1-x)U_{k+1}+xU_k$ we find that the function defined by this expression is the function f_k. □

It is understood that $f_k(x)$ is defined by continuity for $x = 0$ and $x = 1$. An application of l'Hôspitals rule shows that

$$f_k(0) = 2u_k - 1\,,\ f_k(1) = \infty\,. \tag{12}$$

Then we investigate the limiting behaviour of $(f_k)_{k\geq 1}$ for $k \to \infty$.

Lemma 2. *The pointwise limit $f = \lim_{k\to\infty} f_k$ exists in $[0,1]$ and is given by*

$$f(x) = \frac{2\left(-x - \, ln\,(1-x)\right)}{x^2(1+x)}\,;\ 0 < x < 1 \tag{13}$$

with $f(0) = 1$ and $f(1) = \infty$. Alternatively,

$$f(x) = \frac{2}{1+x}\sum_{n=0}^{\infty}\frac{x^n}{n+2}\,;\ 0 \leq x \leq 1\,.^{4} \tag{14}$$

---

[3]For the benefit of the reader we point out that this inequality can be derived rather directly from the *lemma of replacement* developed in [4]. The relevant part of that lemma is the following result: If $f : [0,1] \to \mathbb{R}$ is concave/convex (i.e. concave on $[0,\xi]$, convex on $[\xi,1]$ for some $\xi \in [0,1]$), then, for any $P \in M_+^1(\mathbb{N})$, there exists $k \geq 1$ and a convex combination P_0 of U_{k+1} and U_k such that $F(P_0) \leq F(P)$ with F defined by $F(Q) = \sum_1^\infty f(q_n)$; $Q \in M_+^1(\mathbb{N})$.

[4]or, as a power series in x, $f(x) = 2\sum_0^\infty (-1)^n(1-l_{n+2})x^n$ with $l_n = -\sum_1^n (-1)^k \frac{1}{k}$.

The simple proof, based directly on Lemma 1, is left to the reader. We then investigate some of the properties of f:

Lemma 3. *The function f is convex, $f(0) = 1$, $f(1) = \infty$ and $f'(0) = -\frac{1}{3}$. The real number $x_0 = \mathrm{argmin} f$ is uniquely determined by one of the following equivalent conditions:*

(i) $f'(x_0) = 0$,

(ii) $-\, ln\,(1 - x_0) = \frac{2x_0(1+x_0-x_0^2)}{(3x_0+2)(1-x_0)}$,

(iii) $\sum_{n=1}^{\infty} \left(\frac{n+1}{n+3} + \frac{n-1}{n+2}\right) x_0^n = \frac{1}{6}$.

One has $x_0 \approx 0.2204$ and $\gamma \approx 0.9640$ with $\gamma = f(x_0) = \min f$.

Proof. By standard differentiation, say based on (13), one can evaluate f and f'. One also finds that (i) and (ii) are equivalent. The equivalence with (iii) is based on the expansion

$$f'(x) = \frac{2}{(1+x)^2} \sum_{n=0}^{\infty} \left(\frac{n+1}{n+3} + \frac{n-1}{n+2}\right) x^n$$

which follows readily from (14).

The convexity, even strict, of f follows as f can be written in the form

$$f(x) = \left(\frac{2}{3} + \frac{1}{3} \cdot \frac{1}{1+x}\right) + \sum_{n=2}^{\infty} \frac{2}{n+2} \frac{x^n}{1+x} ,$$

easily recognizable as a sum of two convex functions.

The approximate values of x_0 and γ were obtained numerically, based on the expression in (ii). $\square$

The convergence of f_k to f is in fact increasing:

Lemma 4. *For every $k \geq 1$, $f_k \leq f_{k+1}$.*

Proof. As a more general result will be proved as part (i) of Theorem 2, we only indicate that a direct proof involving three times differentiation of the function

$$\Delta_k(x) = \frac{1}{2} x^2 (1 - x^2)(f_{k+1}(x) - f_k(x))$$

is rather straight forward. $\square$

Lemma 5. $\gamma_1 = \; ln\, 4 - 1 \approx 0.3863$.

Proof. We wish to find the best (largest) constant c such that

$$H(P) \geq \ln 4 \cdot (1 - IC(P)) + 2c\left(IC(P) - \frac{1}{2}\right)(1 - IC(P)) \qquad (15)$$

holds for all $P \in M_+^1(\mathbb{N})$, cf. (11), and know that we only need to worry about distributions $P \in M_+^1(2)$. Let $P = (p, q)$ be such a distribution, i.e. $0 \leq p \leq 1$, $q = 1 - p$. Take p as independent variable and define the auxiliary function $h = h(p)$ by

$$h = H - \ln 4 \cdot (1 - IC) - 2c(IC - \frac{1}{2})(1 - IC)\,.$$

Here, $H = -p \ln p - q \ln q$ and $IC = p^2 + q^2$. Then:

$$h' = \ln \frac{q}{p} + 2(p - q) \ln 4 - 2c(p - q)(3 - 4IC)\,,$$
$$h'' = -\frac{1}{pq} + 4 \ln 4 - 2c(-10 + 48pq)\,.$$

Thus $h(0) = h(\frac{1}{2}) = h(1) = 0$, $h'(0) = \infty$, $h'(\frac{1}{2}) = 0$ and $h'(1) = -\infty$. Further, $h''(\frac{1}{2}) = -4 + 4 \ln 4 - 4c$, hence h assumes negative values if $c > \ln 4 - 1$. Assume now that $c < \ln 4 - 1$. Then $h''(\frac{1}{2}) > 0$. As h has (at most) two inflection points (follows from the formula for h'') we must conclude that $h \geq 0$ (otherwise h would have at least six inflection points!).

Thus $h \geq 0$ if $c < \ln 4 - 1$. Then $h \geq 0$ also holds if $c = \ln 4 - 1$. □

The lemma is an improvement over an inequality established in [8] as we shall comment more on in Section 3.

With relatively little extra effort we can find reasonable bounds for each of the γ_k's in terms of γ. What we need is the following lemma:

Lemma 6. *For $k \geq 1$ and $0 \leq x < 1$,*

$$f_k(x) = \frac{2k}{(k+1)(1-x^2)} \sum_{n=0}^{\infty} \frac{1}{2n+2} \cdot \left[\frac{1 - x^{2n+1}}{2n+3}\left(1 - \frac{1}{k^{2n+2}}\right) + \frac{1 - x^{2n}}{2n+1}\left(1 + \frac{1}{k^{2n+1}}\right)\right] \qquad (16)$$

and

$$f(x) = \frac{2}{1-x^2} \sum_{n=0}^{\infty} \frac{1}{2n+2}\left(\frac{1 - x^{2n+1}}{2n+3} + \frac{1 - x^{2n}}{2n+1}\right)\,. \qquad (17)$$

Proof. Based on the expansions

$$-x - \ln(1-x) = x^2 \sum_{n=0}^{\infty} \frac{x^n}{n+2}$$

and

$$(k+x)\ln(1+\frac{x}{k}) = x + x^2 \sum_{n=0}^{\infty} \frac{(-1)^n x^n}{(n+2)(n+1)k^{n+1}}$$

(which is also used for $k=1$ with x replaced by $-x$), one readily finds that

$$\begin{aligned} &-(k+x)\ln(1+\frac{x}{k}) - (1-x)\ln(1-x) + (k+1)x^2 \ln(1+\frac{1}{k}) \\ &= x^2 \left[1 + \sum_{n=0}^{\infty} \frac{(-1)^n}{(n+2)(n+1)} \cdot \frac{1}{k^{n+1}} - \sum_{n=0}^{\infty} \frac{x^n}{(n+2)(n+1)} \left(\frac{(-1)^n}{k^{n+1}} + 1\right)\right]. \end{aligned}$$

Upon writing 1 in the form

$$1 = \sum_{n=0}^{\infty} \frac{1}{2n+2} \left(\frac{1}{2n+1} + \frac{1}{2n+3}\right)$$

and collecting terms two-by-two, and subsequent division by $1-x^2$ and multiplication by $2k$, (16) emerges. Clearly, (17) follows from (16) by taking the limit as k converges to infinity. □

Putting things together, we can now prove the following result:

Theorem 1. *We have* $\gamma_1 \le \gamma_2 \le \cdots$, $\gamma_1 = \ln 4 - 1 \approx 0.3863$ *and* $\gamma_k \to \gamma$ *where* $\gamma \approx 0.9640$ *can be defined as*

$$\gamma = \min_{0<x<1} \left\{ \frac{2}{x^2(1+x)} \left(\ln \frac{1}{1-x} - x \right) \right\}.$$

Furthermore, for each $k \ge 1$,

$$(1-\frac{1}{k})\gamma \le \gamma_k \le (1 - \frac{1}{k} + \frac{1}{k^2})\gamma. \tag{18}$$

Proof. The first parts follow directly from lemmas 1-5. To prove the last statement, note that, for $n \ge 0$,

$$1 - \frac{1}{k^{2n+2}} \ge 1 - \frac{1}{k^2}.$$

It then follows from Lemma 6 that $(1+\frac{1}{k})f_k \ge \left(1 - \frac{1}{k^2}\right) f$, hence $f_k \ge (1-\frac{1}{k})f$ and and $\gamma_k \ge (1-\frac{1}{k})\gamma$ follows.

Similarly, note that $1+k^{-(2n+1)} \leq 1+k^{-3}$ for $n \geq 1$ (and that, for $n=0$, the second term in the summation in (16) vanishes). Then use Lemma 6 to conclude that $(1+\frac{1}{k})f_k \leq (1+\frac{1}{k^3})f$. The inequality $\gamma_k \leq (1-\frac{1}{k}+\frac{1}{k^2})\gamma$ follows. □

The discussion contains more results, especially, the bounds in (18) are sharpened.

3 Discussion and further results

Justification:

The justification for the study undertaken here is two-fold: As a study of certain aspects of the relationship between entropy and index of coincidence – which is part of the wider theme of comparing one Rényi entropy with another, cf. [4] and [10] – and as a preparation for certain results of exact prediction in Bernoulli trials. The former type of justification was carefulle dealt with in Section 1.

Regarding the latter type of justification, related to prediction, let us briefly indicate what is involved. We consider a Bernouilli source generating a string of symbols $x_1, x_2, x_3, \cdots$ from an n-letter alphabet. For the special model we have in mind, it is assumed that the source distribution is a uniform distribution over a subset of the alphabet – possibly over the entire alphabet. The goal is, given integers $0 \leq s < t$, to make *predictions* of $x_{s+1}^t = x_{s+1} \cdots x_t$ based on knowledge of $x_1^s = x_1 \cdots x_s$. Equivalently, one seeks good *codes* (or *descriptions*) of x_{s+1}^t. Unique optimal objects exist, theoretically, but are difficult, or often even impossible to calculate in closed form. The central technical tool needed in order to make exact calculations in cases where this is possible turns out to be inequalities of the type here studied. Presently it is not possible to point to literature which explains in more detail why this is the case. However, the short proceedings contribution [5] may be helpful in this respect.

Lower bounds for distributions over a two-element alphabet:

Regarding Lemma 5, the key result proved is really the following inequality for a two-element probability distribution $P = (p, q)$:

$$4pq\left(\ln 2 + (\ln 2 - \frac{1}{2})(1-4pq)\right) \leq H(p,q)\,. \tag{19}$$

Let us compare this with the lower bounds contained in the following inequalities proved in [8]:

$$
\begin{aligned}
ln\, p\, ln\, q \leq & H(p,q) \leq \frac{ln\, p\, ln\, q}{ln\, 2}\,, & (20)\\
ln\, 2 \cdot 4pq \leq & H(p,q) \leq\ ln\, 2(4pq)^{1/\, ln\, 4}\,. & (21)
\end{aligned}
$$

Clearly, (19) is sharper than the lower bound in (21). Numerical evidence shows that "normally" (19) is also sharper than the lower bound in (20) but, for distributions close to a deterministic distribution, (20) is in fact the sharper of the two.

More on the convergence of f_k to f:

Though Theorem 1 ought to satisfy most readers, we shall continue and derive sharper bounds than those in (18). This will be achieved via a closer study of the functions f_k and their convergence to f as $k \to \infty$. By looking at previous results, notably perhaps Lemma 1 and the proof of Theorem 1, the suspicion is raised that it is the sequence of functions $(1+\frac{1}{k})f_k$ rather than the sequence of f_k's that are well behaved. This is supported by the results assembled in the theorem below, which, at least for parts (ii) and (iii), are rather combersome to establish:

Theorem 2. *(i)* $(1+\frac{1}{k})f_k \uparrow f$, *i.e.* $2f_1 \leq \frac{3}{2}f_2 \leq \frac{4}{3}f_3 \leq \cdots \to f$.
(ii) For each $k \geq 1$, *the function* $f - (1+\frac{1}{k})f_k$ *is decreasing in* $[0,1]$.
(iii) For each $k \geq 1$, *the function* $(1+\frac{1}{k})f_k/f$ *is increasing in* $[0,1]$.

The technique of proof will be elementary, mainly via turturous differentiations (which may be replaced by cowardice MAPLE look-ups, though) and will rely also on certain inequalities for the logarithmic function in terms of rational functions. The proof is relegated to the appendix.

An analogous result appears to hold for convergence from above to f. Indeed, experiments on MAPLE indicate that $(1+\frac{1}{k}+\frac{1}{k^2})f_k \downarrow f$ and that natural analogs of (ii) and (iii) of Theorem 2 hold. However, this will not lead to improved bounds over those der ived below in Theorem 3.

Refined bounds for γ_k in terms of γ:

Such bounds follow easily from (ii) and (iii) of Theorem 2:

Theorem 3. *For each* $k \geq 1$, *the following inequalities hold:*

$$
(2u_k - 1)\gamma \leq \gamma_k \leq \frac{k}{k+1}\gamma + \frac{2k}{k+1} - \frac{2k+1}{k+1}u_k\,. \qquad (22)
$$

Proof. Define constants a_k and b_k by

$$a_k = \inf_{0 \le x \le 1} \left(f(x) - (1 + \frac{1}{k}) f_k(x) \right) ,$$
$$b_k = \inf_{0 \le x \le 1} \frac{(1 + \frac{1}{k}) f_k(x)}{f(x)} .$$

Then

$$b_k \gamma \le (1 + \frac{1}{k}) \gamma_k \le \gamma - a_k .$$

Now, by (ii) and (iii) of Theorem 2 and by an application of l'Hôpitals rule, we find that

$$a_k = (2 + \frac{1}{k}) u_k - 2 ,$$
$$b_k = (1 + \frac{1}{k})(2u_k - 1) .$$

The inequality (22) follows. □

Note that another set of inequalities can be obtained by working with sup instead of inf in the definitions of a_k and b_k. However, inspection shows that the inequalities obtained that way are weaker than those given by (22).

The inequalities (22) are sharper than (18) of Theorem 1 but less transparent. Simpler bounds can be obtained by exploiting lower bounds for u_k (obtained from lower bounds for *ln* $(1 + x)$, cf. [8]). One such lower bound is given in footnote [2] and leads to the inequalities

$$\frac{2k-1}{2k+1} \gamma \le \gamma_k \le \frac{k}{k+1} \gamma . \tag{23}$$

Of course, the upper bound here is also a consequence of the relatively simple property (i) of Theorem 2. When you apply the bound (27) quoted further on of the logarithmic function, the inequalities in(23) are sharpened as follows:

$$\frac{6k^2 - 1}{6k^2 + 6k + 1} \gamma \le \gamma_k \le \frac{k}{k+1} \left(\gamma - \frac{1}{6k^2 + 6k + 1} \right) . \tag{24}$$

Appendix

We shall here give the proof of Theorem 2. We need some auxiliary bounds for the logarithmic function which were researched in their own right and are available from [9]. In particular, we quote a result for the function λ defined by

$$\lambda(x) = \frac{\mathit{ln}\,(1+x)}{x}\,.$$

What we have in mind is the following double inequality:

$$(2-x)\lambda(y) - \frac{1-x}{1+y} \le \lambda(xy) \le x\lambda(y) + (1-x)\,, \tag{25}$$

cf. (16) of [9]. The inequalities here are valid for $0 \le x \le 1$ and $0 \le y < \infty$ (with $\lambda(0) = 1$).

Proof of (i) of Theorem 2:

Fix $0 \le x \le 1$ and introduce the parameter $y = \frac{1}{k}$. Put

$$\psi(y) = (1+\frac{1}{k})\frac{x^2(1-x^2)}{2} f_k(x) + (1-x)\;\mathit{ln}\,(1-x)$$

(with $k = \frac{1}{y}$). Then

$$\psi(y) = \frac{-(1+xy)\;\mathit{ln}\,(1+xy) + x^2(1+y)\;\mathit{ln}\,(1+y)}{y}\,.$$

Allow y to vary in $]0,1]$. We will show that ψ is a decreasing function of y. This will imply the desired result. We find:

$$\psi'(y) = \frac{1}{y^2}\left(-xy + x^2y + \;\mathit{ln}\,(1+xy) - x^2\;\mathit{ln}\,(1+y)\right)$$

which is ≤ 0 in view of the right hand inequality of (25). As $\psi' \le 0$, ψ is decreasing as claimed.

Proof of (ii) of Theorem 2:

Fix $k \ge 1$ and put $\varphi = f - \left(1+\frac{1}{k}\right) f_k$. Then

$$\varphi(x) = \frac{2}{x^2(1-x^2)}\left[(k+x)\,\mathit{ln}\,\left(1+\frac{x}{k}\right) - x^2(k+1)\,\mathit{ln}\,\left(1+\frac{1}{k}\right) - x + x^2\right]$$

and φ' can be written in the form

$$\varphi'(x) = \frac{2kx}{x^4(1-x^2)^2}\psi(x)$$

where, with $y = \frac{1}{k}$,

$$\begin{aligned}\psi(x) =\,& xy(3x^2-1)\,\mathit{ln}\,(1+xy) + (4x^2-2)\,\mathit{ln}\,(1+xy)\\ &- 2x^4(1+y)\,\mathit{ln}\,(1+y) + 2xy(x^3-2x^2+1)\,.\end{aligned}$$

We have to prove that $\psi \le 0$ in $[0,1]$. One finds that

$$\begin{aligned}\psi'(x) =&(9x^2y - y + 8x)\, ln\,(1+xy) - 8x^3(1+y)\, ln\,(1+y)\\ &+ y(8x^3 - 12x^2 + 2) + \frac{3x^3y^2 - xy^2 + 4x^2y - 2y}{1+xy}\end{aligned}$$

and, further, that

$$\begin{aligned}\psi''(x) =&(18xy+8)\, ln\,(1+xy) - 24x^2(1+y)\, ln\,(1+y)\\ &+ 24x^2y - 9xy + \frac{xy(1-y^2)}{(1+xy)^2}\,.\end{aligned}$$

From the formulas for ψ, ψ' and ψ'' one finds that

$$\psi(0) = \psi(1) = \psi'(0) = \psi'(1) = \psi''(0) = 0\,.$$

Furthermore, we claim that $\psi''(1) < 0$. This amounts to the inequality

$$ln\,(1+y) > \frac{y(8+7y)}{(1+y)(8+3y)}\,. \tag{26}$$

This inequality, which is valid for $y > 0$, may either be proved directly by elementary means or one may deduce it from the stronger inequality

$$ln\,(1+y) > \frac{3y(2+y)}{6+6y+y^2}\,, \tag{27}$$

which, in turn, is known from the theory of Padé approximation, cf. [9] (the right hand side of (27) is the Padé approximant ϕ_2, listed in Table 1 of [9]).

Further differentiation yields

$$\begin{aligned}\psi'''(x) = &\,18y\, ln\,(1+xy) - 48x(1+y)\, ln\,(1+y) + 9y + 48xy\\ &- \frac{10y}{1+xy} - \frac{y(1-y^2)}{(1+xy)^2} + \frac{2y(1-y^2)}{(1+xy)^3}\,,\end{aligned}$$

hence $\psi'''(0) = -y^3 < 0$. We need two more differentiations:

$$\begin{aligned}\psi^{(4)}(x) = &\,\frac{18y^2}{1+xy} - 48(1+y)\, ln\,(1+y) + 48y\\ &+ \frac{10y^2}{(1+xy)^2} + \frac{2y^2(1-y^2)}{(1+xy)^3} - \frac{6y^2(1-y^2)}{(1+xy)^4}\,,\\ \psi^{(5)}(x) = &- \frac{18y^3}{(1+xy)^2} - \frac{20y^3}{(1+xy)^3} - \frac{6y^3(1-y^2)}{(1+xy)^4} + \frac{24y^3(1-y^2)}{(1+xy)^5}\,.\end{aligned}$$

Now, if ψ assumes positive values in $[0,1]$, $\psi''(x)=0$ would have at least 4 solutions in $]0,1[$. Then $\psi'''(x)=0$ would have at least 3, $\psi^{(4)}(x)=0$ at least 2 and $\psi^{(5)}(x)=0$ at least one solution in $]0,1[$. In order to arrive at a contradiction, we put $X=1+xy$ and note that $\psi^{(5)}(x)=0$ is equivalent to the equality

$$-9X^3-10X^2-3(1-y^2)X+12(1-y^2)=0\,.$$

Now, the left hand side here is upper bounded by

$$-9-10-3(1-y^2)+12(1-y^2)=-10-9y^2\,,$$

a negative number. This gives us the desired contradiction, hence $\psi\leq 0$ in $[0,1]$.

Proof of (iii) of Theorem 2:

Again, fix k and put

$$\psi(x)=1-\frac{(1+\frac{1}{k})f_k(x)}{f(x)}\,.$$

Then, once more with $y=\frac{1}{k}$,

$$\psi(x)=\frac{(1+xy)\,\ln\,(1+xy)-x^2(1+y)\,\ln\,(1+y)-xy(1-x)}{y(1-x)(-x-\,\ln\,(1-x))}\,.$$

We will show that $\psi'\leq 0$. Write ψ' in the form

$$\psi'=\frac{y}{\text{denominator}^2}\xi\,,$$

where "denominator" refers to the denominator in the expression for ψ. Then

$$\begin{aligned}\xi(x)&=(y\,\ln\,(1+xy)-(2x+2xy)\,\ln\,(1+y)+2xy)\big(-x+x^2-(1-x)\,\ln\,(1-x)\big)\\&\quad-\big((1+xy)\,\ln\,(1+xy)-x^2(1+y)\,\ln\,(1+y)-xy+x^2y\big)\,(2x+\,\ln\,(1-x))\\&=-(1+y)\,\ln\,(1+xy)\,\ln\,(1-x)+x(2-x)(1+y)\,\ln\,(1+y)\,\ln\,(1-x)\\&\quad-x(xy+y+2)\,\ln\,(1+xy)+2x^2(1+y)\,\ln\,(1+y)-2xy(1-x)^2\,.\end{aligned}$$

It follows that $\xi(0)=\xi_k(1)=0$. In more detail regarding the continuity of ξ at 1 with $\xi(1)=0$, the key fact needed is the limit relation

$$\lim_{x-1^-}\,\ln\,(1-x)\cdot\,\ln\,\frac{1+xy}{1+y}=0\,.$$

This follows easily from the general inequality $ln\, x \leq \frac{x-1}{\sqrt{x}}$ for $x \geq 1$, cf. (14) of [9].

We differentiate:

$$\begin{aligned}
\xi'(x) &= -(1+y)\left(\frac{y}{1+xy}\ ln\ (1-x) - \frac{1}{1-x}(1+xy)\right) \\
&\quad + (1+y)\ ln\ (1+y) \cdot \left((2-2x)\ ln\ (1-x) - x(2-x)\cdot\frac{1}{1-x}\right) \\
&\quad - \left((2xy+y+2)\ ln\ (1+xy) + x(xy+y+2)\frac{y}{1+xy}\right) \\
&\quad + 4x(1+y)\ ln\ (1+y) - 2y(1-x)^2 + 4xy(1-x)\,.
\end{aligned}$$

Thus $\xi'(0) = -2y < 0$, $\xi'(1) = \infty$. Further differentiation and exploitation of the left hand inequality of (25) gives:

$$\begin{aligned}
\xi''(x) &= \left(2(1+y)\ ln\ (1+y) - \frac{(1+y)y^2}{(1+xy)^2}\right)\cdot(-\ ln\ (1-x)) \\
&\quad + \left(\frac{1+y}{(1-x)^2} - 2y\right)\ ln\ (1+xy) - (1+y)\frac{x(2-x)}{(1-x)^2}\ ln\ (1+y) \\
&\quad + y\left(5 + \frac{2}{1-x} - \frac{1-xy^2}{(1+xy)^2} - 12x\right) \\
&\geq \left(2(1+y)\ ln\ (1+y) - \frac{(1+y)y^2}{(1+xy)^2}\right)(-\ ln\ (1-x)) \\
&\quad - 2y\ ln\ (1+xy) + y\left(5 + \frac{2-x}{1-x} - \frac{1-xy^2}{(1+xy)^2} - 12x\right) \\
&\geq \left(2y - \frac{(1+y)y^2}{(1+xy)^2}\right)x - 2y^2x + y\left(6 + \frac{1}{1-x} - \frac{1-xy^2}{(1+xy)^2} - 12x\right) \\
&= -10xy - 2y^2x - \frac{y}{1+xy} + 6y + \frac{y}{1-x}\,.
\end{aligned}$$

Dividing by y we obtain the expression

$$-10x - 2xy - \frac{1}{1+xy} + 6 + \frac{1}{1-x}\,,$$

which we have to prove is ≥ 0. Now note that, for fixed x,

$$-2xy - \frac{1}{1+xy} \geq -2x - \frac{1}{1+x}\,,$$

thus we have to prove that $\zeta \geq 0$ with

$$\zeta(x) = -12x - \frac{1}{1+x} + \frac{1}{1-x} + 6\,.$$

By simple differentiation we find that ζ is minimal for $x = \frac{1}{2}\sqrt{2}$ with minimum $6 - 4\sqrt{2} \approx 0.34 > 0$.

All parts of Theorem 2 are hereby proved.

References

[1] M. Ben-Bassat. f-entropies, probability of error, and feature selection. *Information and Control*, 39:227–242, 1978.

[2] M. Feder and N. Merhav. Relations between entropy and error probability. *IEEE Trans. Inform. Theory*, 40:259–266, 1994.

[3] J. DJ. Golić. On the relationship between the information measures and the Bayes probability of error. *IEEE Trans. Inform. Theory*, IT-33(5):681–690, Sept. 1987.

[4] P. Harremoës and F. Topsøe. Inequalities between Entropy and Index of Coincidence derived from Information Diagrams. *IEEE Trans. Inform. Theory*, 47(7):2944–2960, Nov. 2001.

[5] P. Harremoës and F. Topsøe. Unified approach to Optimization Techniques in Shannon Theory. In *Proceedings, 2002 IEEE International Symposium on Information Theory*, page 238. IEEE, 2002.

[6] V. A. Kovalevskij. *The problem of character recognition from the point of view of mathematical statistics*, pages 3–30. Spartan, New York, 1967.

[7] D.L. Tebbe and S.J. Dwyer. Uncertainty and the probability of error. *IEEE Trans. Inform. Theory*, 14:516–518, 1968.

[8] F. Topsøe. Bounds for entropy and divergence of distributions over a two-element set. *J. Ineq. Pure Appl. Math.*, 2001. [ONLINE] http://jipam.vu.edu.au/accepted_papers/ 044_00.html.

[9] F. Topsøe. Some bounds for the logarithmic function. To appear, available from the RGMIA preprint server, see http://rgmia.vu.edu.au/v7n2.html, 2004.

[10] K. Zyczkowski. Rényi Extrapolation of Shannon Entropy. *Open Systems and Information Dynamics*, 10:297–310, 2003.